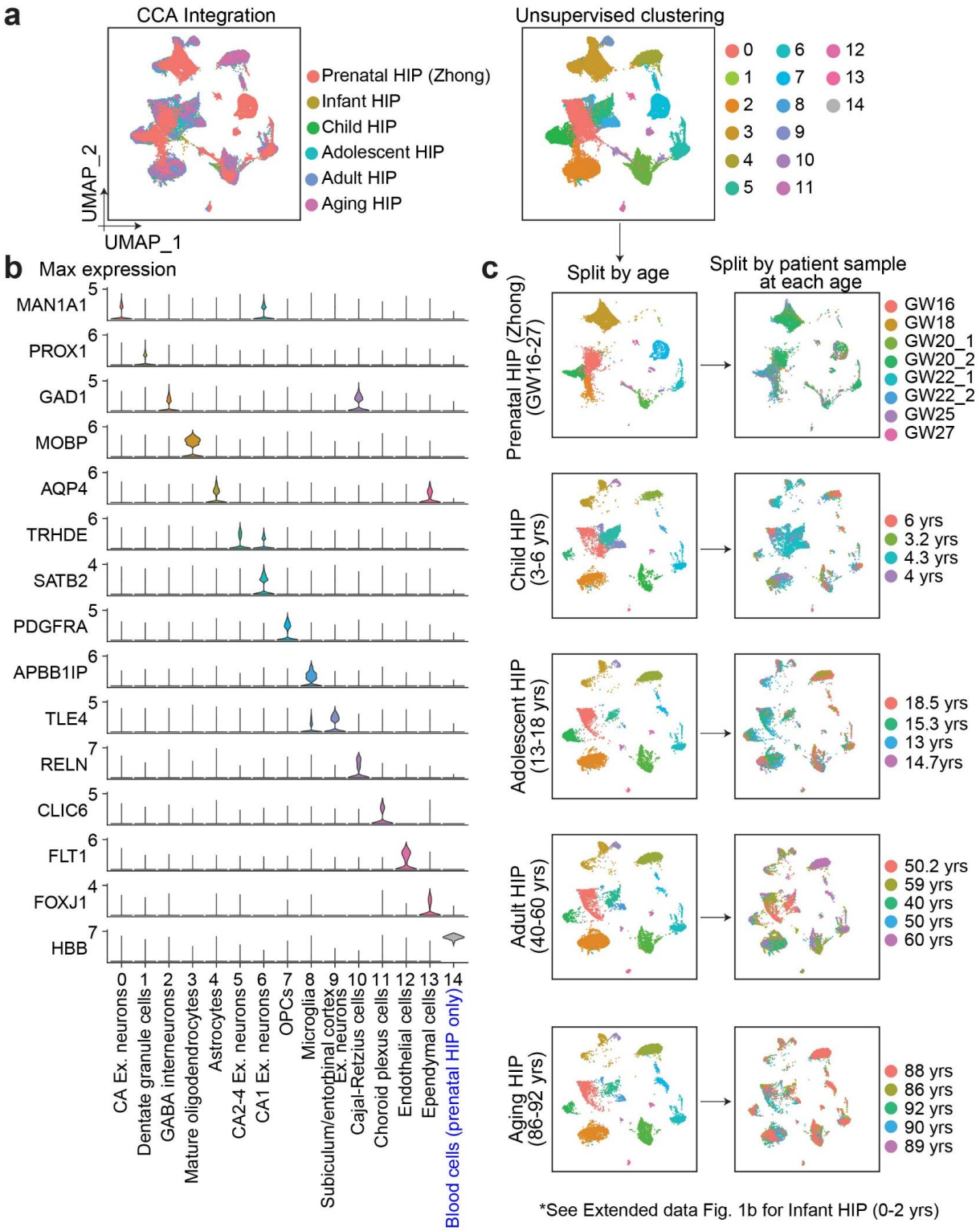
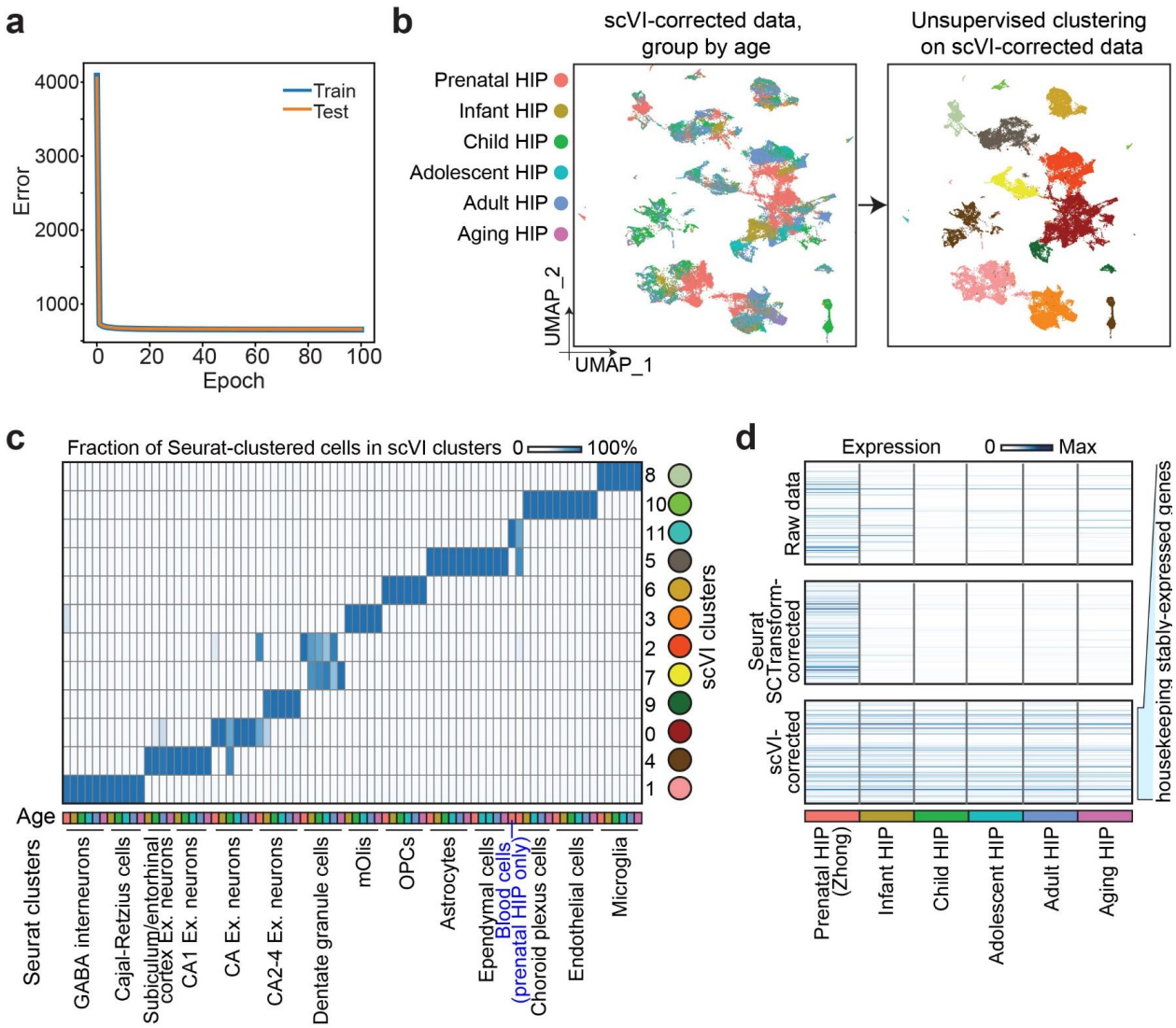
# Molecular landscapes of human hippocampal immature neurons across lifespan

## Supplementary Figures



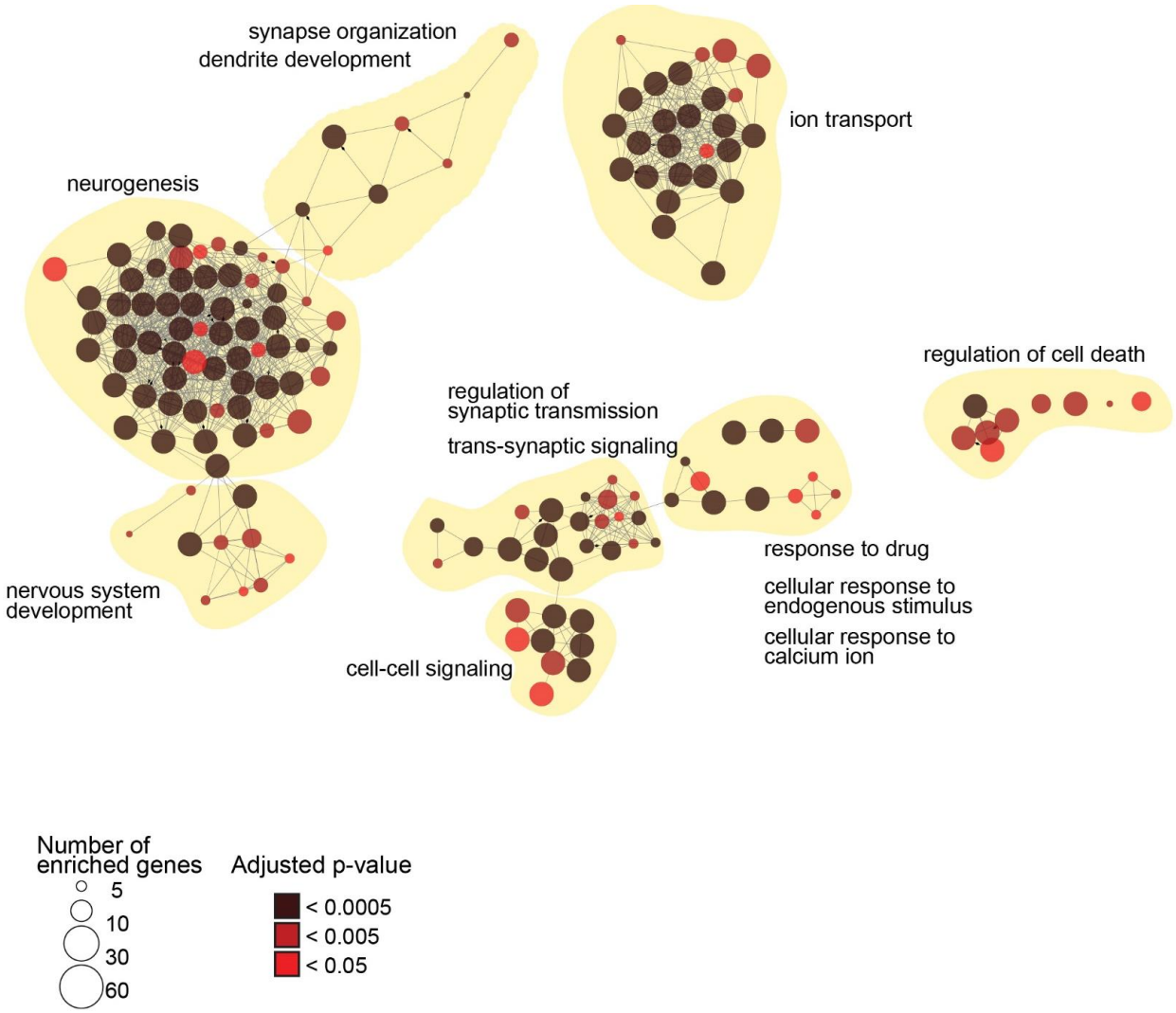

**Supplementary Fig. 1 | Integration and characteristics of human hippocampal scRNA-seq/snRNA-seq datasets across the lifespan. a**, UMAP visualization of cells from human

hippocampal scRNA-seq/snRNA-seq datasets across the lifespan upon integration using canonical correlation analysis (CCA)[36], colored by age (left) and cluster (right). The prenatal dataset was from a published study[34] and all postnatal ones were generated in the current study. **b**, Expression patterns of marker genes used to determine cluster identities. **c**, UMAP visualization of the integrated human hippocampal dataset split by age (left column) and by specimen within each age group (right column). GW: gestational week; yrs: years. See Fig. 1b and Extended Data Fig. 1b for plots for the infant group (0-2 years). See Supplementary Table 1 for de-identified specimen information, Supplementary Table 2 for sequencing characteristics, and Supplementary Table 3 for information on the previously published datasets that were used.

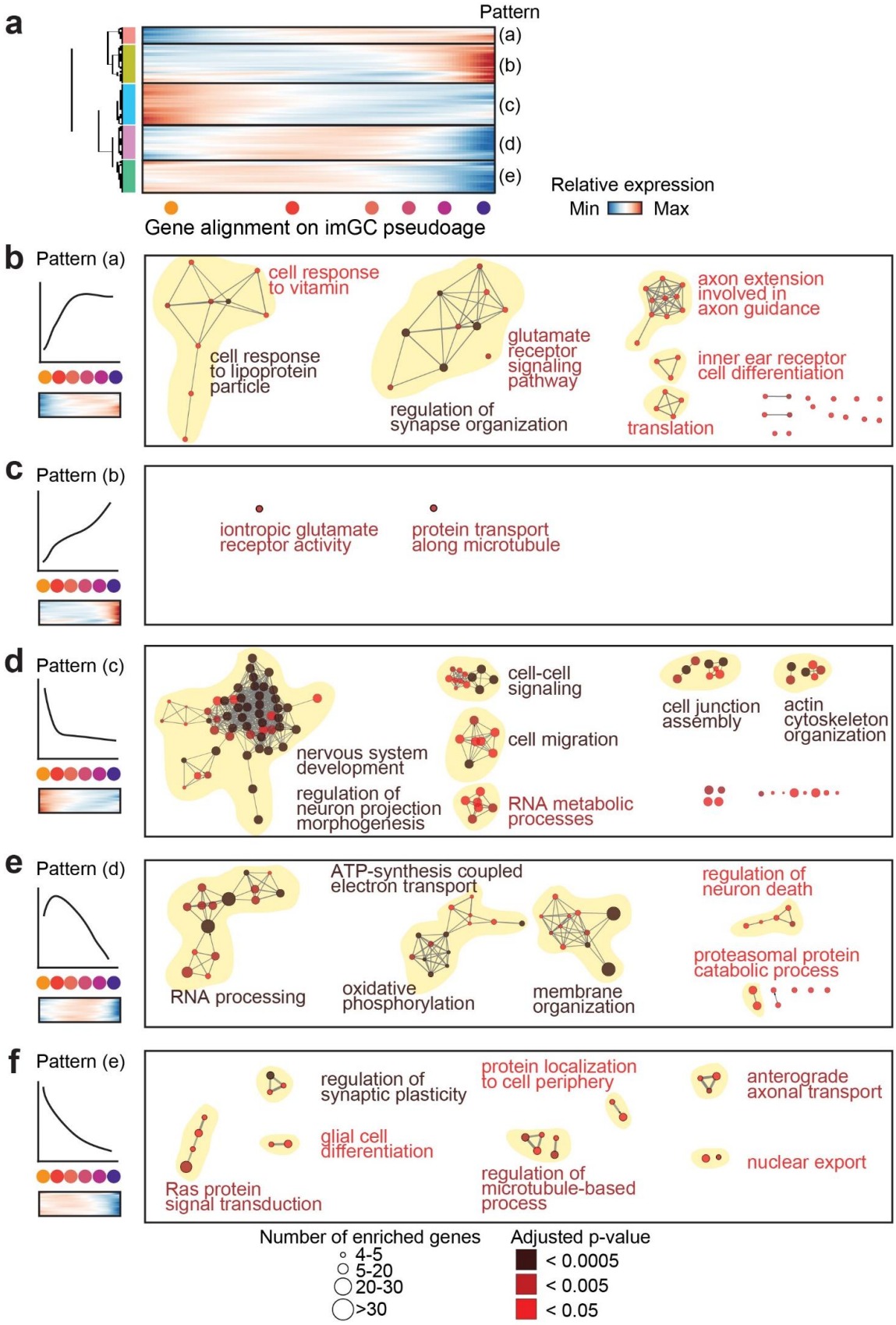



**Supplementary Fig. 2 | scVI correction of hippocampal scRNA-seq/snRNA-seq datasets across ages for sequencing variation. a**, Performance of the scVI algorithm[40], a deep generative modeling analytic method for correcting scRNA-seq/snRNA-seq data matrices for batch effect removal. Line plot showing the likelihood change for the training error (blue) and the testing error (orange) across the 100 epochs of training. **b**, UMAP visualization of cells from human hippocampal scRNA-seq/snRNA-seq datasets across ages after scVI correction and unsupervised clustering, colored by age (left) and by cluster (right). **c**, Heatmap showing the overlap of cluster membership using the dataset processed by scVI (y-axis) versus that aligned with canonical correlation analysis (x-axis,

CCA[36]). Colors indicate the fraction of total cells per CCA-aligned cluster per age group assigned to each scVI cluster. Robust clustering was achieved post-scVI correction with excellent cluster correspondence to the results from CCA, a state-of-the-art cell alignment tool, indicating effective batch correction. **d**, Heatmap showing the expression of the housekeeping "stably-expressed genes"[39,63] in all cells across ages in the uncorrected raw dataset versus after the correction and integration using the 'SC Transform' (built in Seurat[59]) and the scVI methods.

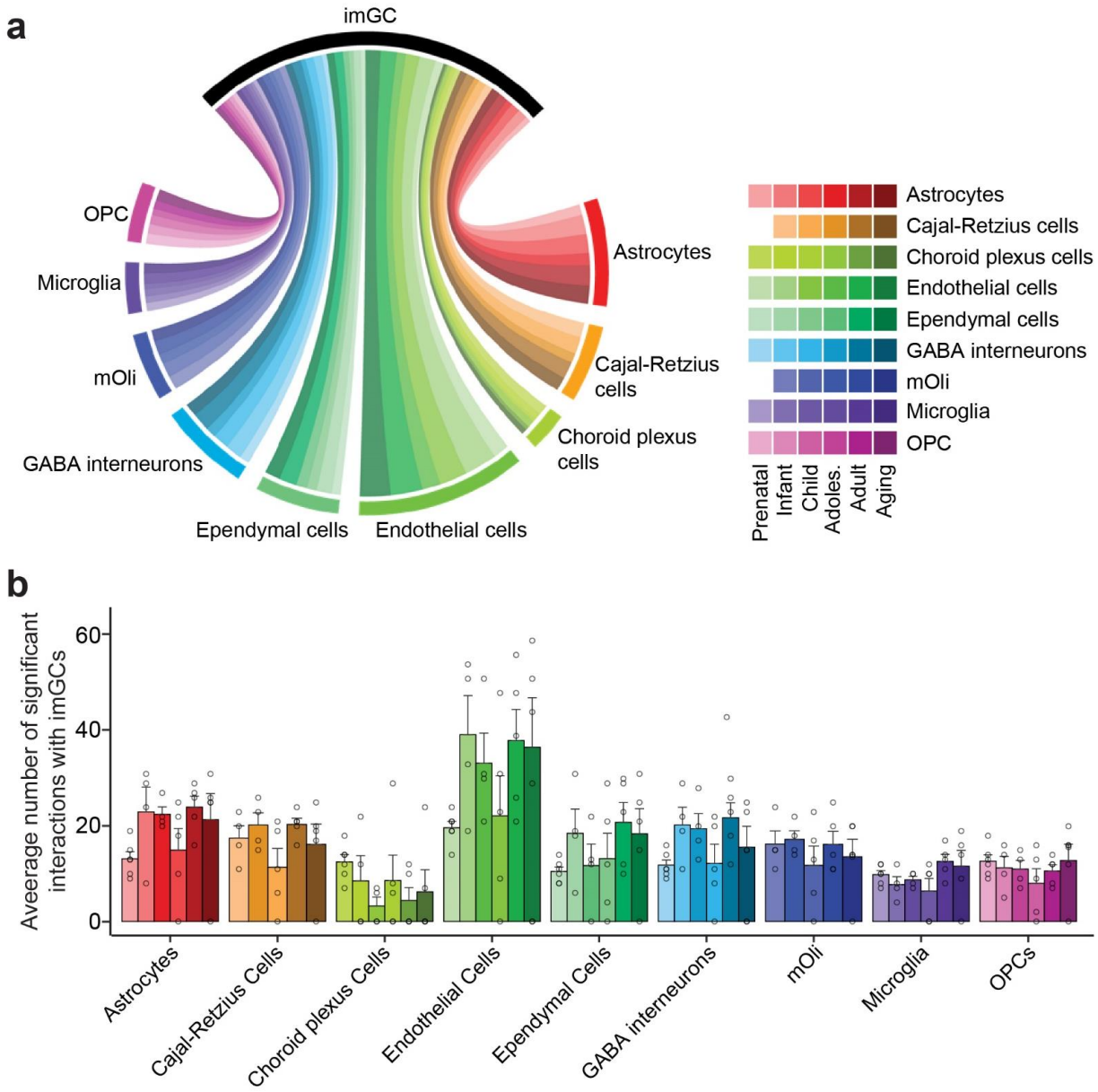



**Supplementary Fig. 3 | Common molecular signatures of human imGCs irrespective of age.** Gene ontology network of biological processes associated with the common human imGCs-enriched genes across ages in comparison to mGCs, colored by FDR-adjusted p-value. Only significantly enriched nodes are displayed (one-sided hypergeometric test, p(FDR) < 0.05). The node size represents the term enrichment significance. Examples of the most significant terms per group are shown. See Fig. 3b for a summary bar plot and Supplementary Table 6 for the list of GO terms.

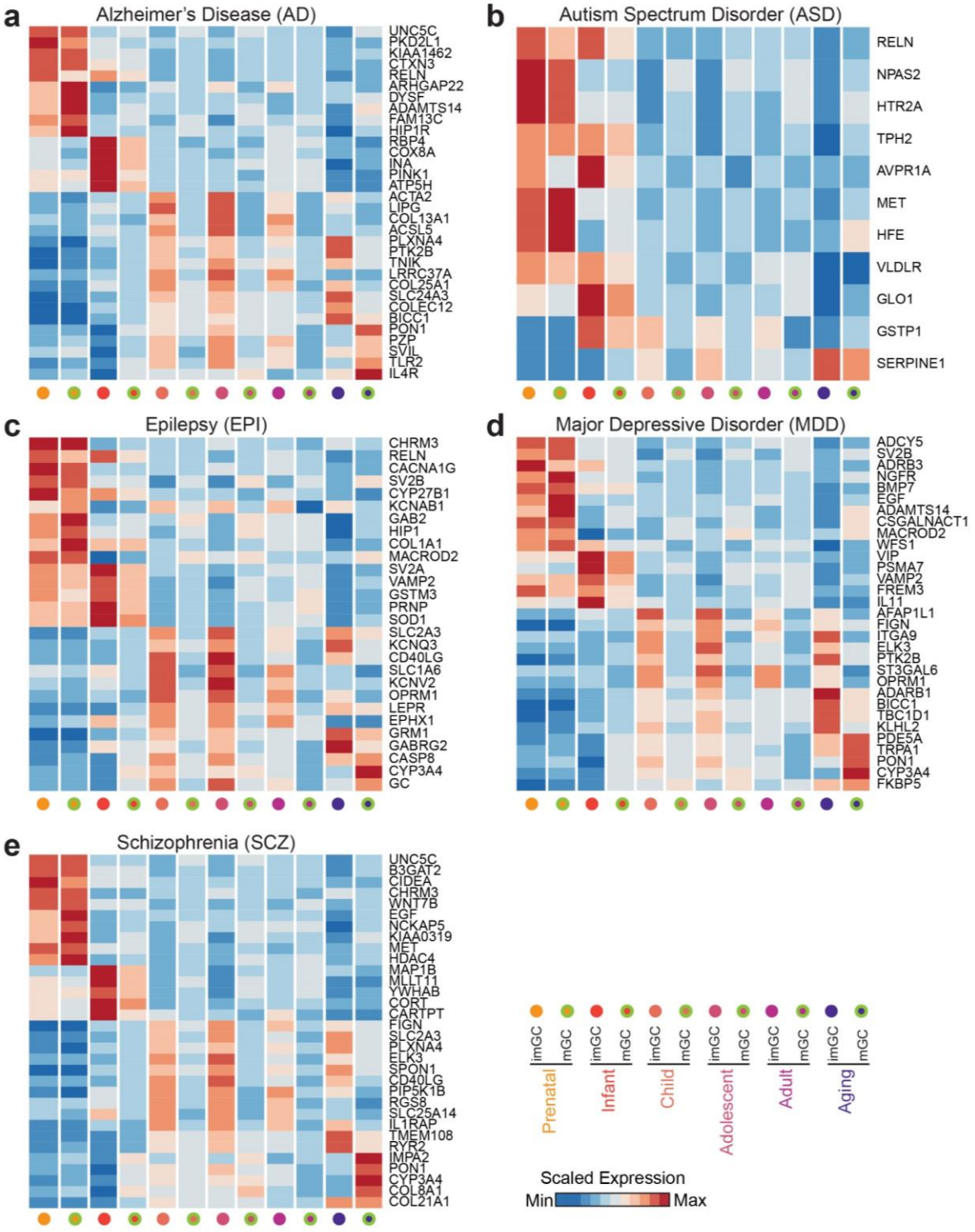



**Supplementary Fig. 4 | Transcriptomic shifts in human imGCs across ages. a,** Pseudo-age gene

co-variation kinetics analysis in Monocle[41] revealed five distinct age-dependent gene expression patterns in human imGCs (likelihood ratio test, Benjamini-Hochberg-adjusted p-value < 0.01, q-value < 0.01). **b-f,** Gene ontology network of biological processes associated with genes in each pattern, colored by FDR-adjusted p-value. Only significantly enriched nodes are displayed (one-sided hypergeometric test, p(FDR) < 0.05). The node size represents the term enrichment significance. Examples of the most significant terms per group are shown. See Supplementary Table 8 for the lists of GO terms.




**Supplementary Fig. 5 | Interactions between human imGCs and different neighboring cell types in the dentate gyrus across ages.** Shown are *Circos* plot (**a**) and bar plot (**b**) displaying the number of specific ligand-receptor interaction pairs between human imGCs and their neighboring cell types in the dentate gyrus of the corresponding specimen across ages. Specificity of the interactions was determined by a one-sided randomization test in CellPhoneDB[42] and a p-value < 0.05 was considered statistically significant. Colors represent cell types within each age group. Dots represent number of significant interactions for each cell type pair in each specimen. Values represent mean ± s.e.m. (n = 28 specimens).




**Supplementary Fig. 6 | Expression of risk genes for neurological disorders in human imGCs and mGCs across the lifespan.** Red-blue heatmaps depict expression patterns of the risk genes of neurological or psychiatric disorders in the human imGCs and mGCs across the lifespan.

## References


34 Zhong, S. *et al.* Decoding the development of the human hippocampus. *Nature* **577**, 531-536, doi:10.1038/s41586-019-1917-5 (2020).

36 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420, doi:10.1038/nbt.4096 (2018).

39 Hodge, R. D. *et al.* Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61-68, doi:10.1038/s41586-019-1506-7 (2019).

40 Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nature Methods* **15**, 1053-+, doi:10.1038/s41592-018-0229-2 (2018).

41 Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**, 979-982, doi:10.1038/nmeth.4402 (2017).

42 Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* **15**, 1484-1506, doi:10.1038/s41596-020-0292-x (2020).

59 Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**, 296, doi:10.1186/s13059-019-1874-1 (2019).

63 Lin, Y. *et al.* Evaluating stably expressed genes in single cells. *Gigascience* **8**, doi:10.1093/gigascience/giz106 (2019).